Beutnagel 3-12-9

REMARKS

Claims 32-34 were rejected under 35 USC 112, first paragraph because one instance of the word "tuple" was inadvertently not replaced. Claim 32 is amended to correct the error and, as amended, it is believed that the rejection is overcome.

Claim 23 was rejected under 35 USC 112, second paragraph because, according to the Examiner the limitation "of said collections" is unclear. Applicants respectfully traverse. The Examiner is apparently confused by the fact that the second bullet item in the claim specifies "an indication of number of parameter information collections, N," and the third bullet item specifies "N parameter information collections." It is suspected that the Examiner's confusion may be the result of parsing the claim incorrectly, and ignoring the comma before the "N," and it is believed that a careful reading of the claim reveals that there is no ambiguity, and no confusion should exist. The second bullet identifies that a number is included (in the detail specification associated with each phoneme). That number is N. That second bullet also specifies that this number corresponds to a number of information collections, and the third bullet specifies those N parameter information collections are included (in the detail specification). Since the first phrase of concern to the Examiner specifies a number, and the second phrase of concern to the Examiner specifies collections of information, no lack of clarity exits when referring to "said collections."

No deficiencies were identified by the Examiner in dependent claims 24-31 and, therefore, in light of the above remarks applicants respectfully submit that the rejection of claims 23-31 has been overcome.

Claims 1-5, 7, 10, 13-22 were rejected under 35 USC 103 as being unpatentable over Yang et al, US Patent 5,970,459 in view of Campbell et al, US Patent 6,366,883. Applicants respectfully traverse.

Claim 1:

The Examiner asserts that Yang et al teach the steps of

(1) inserting a plurality of phonemes (assertion supported by col. 1, lines 35-40);

(2) inserting duration specifications for the phonemes (assertion supported by col. 4, lines 60-66); and

Beutnagel 3-12-9

(3) "including at least one of said phonemes a time offset from the beginning of the duration of said phoneme that is greater than zero and less than the duration of said phone" (assertion supported by col. 5, lines 1-12); and

(4) "at least two prosody parameter specification toward a target value" (assertion supported by col. 4, lines 60-67.

Applicants respectfully disagree.

First, relative to Examiner's assertion (3), applicants note that the passage cited in support of the assertion states:

> The synchronization adjusting unit 14 receives the processing results from the prosody processing unit 13, and adjusts the time durations for every phoneme to synchronize the image signal by using the synchronization information which was received from the multi-media distributor 11. With the adjustment of the time duration of phonemes, the lip shape can be allocated to each phoneme in accordance with the position and manner of articulation for each phoneme, and the series of phonemes is divided into small groups corresponding to the number of the lip shapes recorded in the synchronization information by comparing the lip shape allocated to each phoneme with the lip shape in the synchronization information.

This passage does not mention **any** offsets. It only teaches an adjustment of time durations for every phoneme to synchronize the image signal with the lip shapes. Thus, the passage cited by the Examiner does NOT teach that which the Examiner asserts. Second, applicants note that claim 1 does not specify offsets. To the extent that the time at which a specified parameter reaches a target value is considered an "offset," applicants respectfully direct the Examiner's attention to the discussion below. It might be reiterated, however, that Yang et al do not teach offsets in the sense of applicants' claims and, in fact, the word "offset" is not even found in the Yang et al reference.

Focusing on assertion (4), applicants note that the passage cited in support of the assertion states:

> The prosody processing unit 13 receives the processing results from the language processing unit 12, and calculates the values of the prosodic control parameters. The prosodic control parameter includes the time duration of phonemes, contour of pitch, contour of energy, position of pause, and length. The calculated results are transferred to the synchronization adjusting unit 15.

Beutnagel 3-12-9

This passage teaches that unit 12 calculates values of prosodic control parameters, which parameters are: (1) time duration of the phonemes, (2) contour of pitch, (3) contour of energy, and (4) position and length of pauses. Respectfully, this passage **does not** teach a "target value" specification.

It should be kept in mind that claim 1 uses the word "target," and the word "target" has a specific meaning; defined in the dictionary as "anything aimed or fired at." What that means is that whatever eventually is **at** the target, it certainly is not at the target at an earlier time. One must aim at a target. Thus, one might have a bullet that starts in the barrel of a gun and upon firing later reaches the *target object*, a rocket that starts at zero velocity and later reaches the *target velocity*, a signal that starts not at some unknown voltage level and later reaches the *target voltage value*, etc. Inherently, the notion is of not being at some goal, taking action to reach the goal, and eventually (if successful) reaching that goal after some elapsed time.

Thus, the three words of importance in the claim (target, reaching, and time) are part of one cohesive notion. What the claim specifies is that a prosody parameter is **specified with** two attributes that encompass the three words of importance: a target value, and a time for reaching the target value. No such specification is described or suggested in Yang et al.

To review, what Yang et al teach is specification of a phoneme, and associated with each phoneme there is a specification of either a phoneme, or a pause. Both have a duration specification and phonemes have, additionally, a pitch contour specification and an energy specification. Thus, illustratively, (focusing on phonemes) a Yang et al embodiment may have the string

e,80,P-contour 2,E-contour 17; **o**, 150, P-contour 9,E-contour 4. (1)

To review what claim 1 specifies, in contradistinction, in the first step it specifies inserting a plurality of phonemes represented by symbols; for example, phoneme "e" and "o," or the string

e; **o**.

In the second step claim 1 specifies inserting a signal duration associated with each of the phonemes. The consequence of these two steps, for example, is the string

e,80; **o**,150.

Beutnagel 3-12-9

In the third step claim 1 specifies inserting, in connection with at least one of the phonemes, at least 2 prosody parameter specifications. The consequence of this additional limitation is, for example, the string

e,80; o,150,SPEC1,SPEC2.

The third step also specifies that each specification has a target value and a point in time for reaching this target, and that this point in time "follows beginning of the phoneme and precedes end of the phoneme, unrestricted to any particular point within said duration." The consequence of this limitation is, for example, the string

e,80; o,150, P117@91,P112@141. (2)

Note that both time 91 of the first specification and time 141 of the second specification are both less than 150 – which is the duration of the phoneme.

Comparing string (1) to string (2) it is clear that the specification of a pitch contour and energy contour are *qualitatively* different from the specification of a value and a time, not to mention the specific meaning of the value being a target value an the time being the point in time when the target value is to be reached. **Put very simply, Yang et al do not teach or suggest a string as illustrated in expression (2) above.**

Actually, the Examiner admits that Yang et al fail to "explicitly teach any selected point in time for reaching said target value," but points to a passage at col. 16, lines 14-col. 17, line 23 of the Campbell et al reference that allegedly teaches "a selected point in time for reaching the target value."

Applicants respectfully disagree.

Because the passage cited by the Examiner is quite lengthy, the following presents that passage in table form, with passage texts in the left cells and an explanation of the corresponding passage texts in the right cells.

At step S14, the start position and end position in the speech waveform database file composed of either a plurality of sentences or one sentence for each phoneme segment are recorded, and an index number is assigned to the file. Next, at step S15, the first acoustic feature parameters for each phoneme segment are extracted by using, for example, a known pitch extraction method. Then, at step S16, the phoneme labeling is executed for each phoneme segment, and the phoneme labels and the	This paragraph speaks of feature parameters, but it does NOT speak of values of the parameters (targets or otherwise). The only reference to duration is in the last sentence where it is mentioned that the start positions and durations of

Beutnagel 3-12-9

first acoustic feature parameters for the phoneme labels are recorded. Further, at step S17, the first acoustic feature parameters for each phoneme segment, the phoneme labels and the first prosodic feature parameters for the phoneme labels are stored in the feature parameter memory 30 together with the file index number and the start position and time duration in the file. Finally, at step S18, index information including the index number of the file and the start position and time duration in the file are given to each phoneme segment, and the index information is stored in the feature parameter memory 30, then the speech analysis process is completed.	phonemes are given. The start positions of phonemes are understood to be the start positions of the phonemes in the sequence of phonemes that combine to form an utterance (e.g. a sentence). Relative to the example above, this is akin to the sequence "e,80;o,150."
FIGS. 5 and 6 are flowcharts of the weighting coefficient training process which is executed by the weighting coefficient training controller of FIG. 1.	This paragraph says nothing of parameters, values, times or durations.
Referring to FIG. 5, first of all, at step S21, one phonemic kind is selected from the feature parameter memory 30. Next, at step S22, the second acoustic feature parameters are extracted from the first acoustic feature parameters of a phoneme that has the same phonemic kind as the selected phonemic kind, and then, are taken as the second acoustic feature parameters of the target phoneme. [redacted] At step S24, it is decided whether or not the processes of steps S22 and S23 have been done on all the remaining phonemes. At step S24, if the processes have not been completed for all the remaining phonemes, another remaining phoneme is selected at step S25, and then, the processes of step S23 and the following thereto are iterated.	The notion of a "target" is found in the highlighted (gray) sentence, **but addresses a target phoneme, and not a target value of a parameter of a phoneme.** A target phoneme is the phoneme that one ought to select.
On the other hand, if the processing has been completed at step S24, the top N1 best phoneme candidates are selected at step S26 based on the distances and time durations obtained at step S23. Subsequently, at step S27, the selected N1 best phoneme candidates are ranked into the first to N1-th places. Then, at step S28, for the ranked N1 best phoneme candidates, the scale conversion values are calculated by subtracting intermediate values from the respective distances. Further, at step S29, it is decided whether or not the processes of steps S22 to S28 has been completed for all the phonemic kinds and phonemes. If the processes of steps S22 to S28 have not been completed for all the phonemic kinds, another phonemic kind and	This paragraph discusses selecting "best phoneme candidates" based on distances and time durations; i.e., selecting the target phoneme. This has nothing to do with target **values** of a phoneme's **parameter**. This also has nothing to do with **time to reach the target values.**

Beutnagel 3-12-9

phoneme is selected at step S30, and then the processes of step S22, and the following are iterated. On the other hand, if the processes of steps S22 to S28 has been completed for all the phonemic kinds at step S29, the program flow goes to step S31 of FIG. 6.	
Referring to FIG. 6, at step S31, one phonemic kind is selected. Subsequently, at step S32, the second acoustic feature parameters for each phoneme are extracted for the selected phonemic kind. Then, at step S33, by performing the linear regression analysis based on the scale conversion value for the selected phonemic kind, the degrees of contribution to the scale conversion values in the second acoustic feature parameters are calculated, and the calculated degrees of contribution are stored in the weighting coefficient vector memory 31 as weighting coefficients for each target phoneme. At step S34, it is decided whether or not the processes of steps S32 and S33 has been completed for all the phonemic kinds. If the processes have not been completed for all the phonemic kinds at step S34, another phonemic kind is selected at step S35, and the processes of step S32 and the following are iterated. On the other hand, if the processes has been completed for all the phonemic kinds at step S34, the weighting coefficient training process is completed.	This paragraph speaks of extracting feature parameters and performing regression analysis. There is no mention here of target values, or of times for reaching these target values.

Thus, as the above analysis clearly demonstrates, Campbell et al teach the notion of a target phoneme, but do not teach the notion of a parameter value target. Certainly, Campbell et al do not teach setting a parameter value target AND a point in time when the target is to be reached, and even more certainly, Campbell et al do not teach specifying – in association with a phoneme, at least two specifications that are of like nature, each of which specifying a parameter value as a target to be reached and a point in time when that target value is to be reached.

Thus, applicants respectfully submit that neither Yang et al or Campbell et al individually, nor Yang et al and Campbell et al in combination teach or suggest the limitations of claim 1. Therefore, applicants believed that claim 1 is not obvious in view of the Yang et la and Campbell et al combination of references.

Remaining Claims:

Claims 2-5, 7 and 10-20 depend on claim 1 and therefore are believed to not be obvious in view of the Yang et al and Campbell et al combination of references, at least

by virtue of this dependence. Additionally, it is believed that at least some of the claims contain limitations that make the claims patentable over the Yang et al and Campbell et al combination of references.

Amended claim 2 specifies that at least one phoneme has a specification that includes at least two parameter specifications that BOTH specify pitch. No such notion exits in Yang et al, in col. 7, line 65 (cited by the Examiner) or elsewhere in the reference.

Claim 7 specifies the time at which a parameter reaches its target more particularly, and as indicated above, the entire notion of a parameter reaching a target (i.e. starting at some value other than the target, and traversing some path that eventually makes the parameter have the target value at a given time) is simply not present in either the Yang et al reference or the Campbell et al reference.

Although independent claim 21 was rejected in the group of claims identified in item 4 of the Office action, no explicit comments are offered by the Examiner to justify the rejection.

Amended claim 21 is believed not obvious in view of the Yang et al and the Campbell et al combination of references for the reasons set forth above. Additionally, amended claim 21 explicitly limits the claim to specifications where at least one phoneme has at least one specification that consists of a target value, a time offset, and a delimiter therebetween. No notion of such form to the specification of a control parameter is found in or suggested by either of the cited references. Applicants note that the issue of offset is addressed in the initial treatment of the Examiner's assertions.

As for claim 22, the Examiner has also not provided an explicit comment that explains the reason for the rejection. Applicants note, however, that claim 22 is dependent on claim 21, which is believed to not be obvious in view of the cited references. Moreover, there is no teaching anywhere in the references that the value of a parameter is not restricted, except at the specified offset time. In contradistinction, the contour of pitch and the contour of energy are clearly specified throughout the phoneme's duration (by virtue of the definition of a "contour"), and a pause is also clearly defined throughout – i.e., it being a pause. Therefore, it is respectfully submitted that claim 22 is not obvious in view of the Yang et al and the Campbell et al combination of references.

Beutnagel 3-12-9

In light of the above amendments and remarks, applicants respectfully submit that all of the Examiner's rejections have been overcome. Reconsideration and allowance are respectfully solicited.

Respectfully,
Mark Beutnagel
Joern Ostermann
Schuyler Quackenbusch

Dated: 1/16/06

By _____
Henry T. Brendzel
Reg. No. 26,844
Phone (973) 467-2025
Fax (973) 467-6589
email brendzel@comcast.net